Correcting biased observation model error in data assimilation

Tyrus Berry
Dept. of Mathematical Sciences, GMU

PSU-UMD DA Workshop
June 27, 2017

Joint work with John Harlim, PSU
Consider the standard filtering problem,

\[ x_i = f(x_{i-1}) + \omega_{i-1} \]
\[ y_i = h(x_i) + \eta_i \]

We assume the true observation function \( h(x) \) is unknown.

An approximate model is available \( \tilde{h}(x) \) so that

\[ y_i = h(x_i) + \eta_i = \tilde{h}(x_i) + b_i + \eta_i \]

Where \( b_i \equiv h(x_i) - \tilde{h}(x_i) \) is called the bias.
EXAMPLE 1: LORENZ-96

Consider the standard 40-dimensional Lorenz-96,

\[ \dot{x}_j = x_{j-1}(x_{j+1} - x_{j-2}) - x_j + 8 \]

We observe 20 of the 40 variables

We draw \( \xi_i \sim \mathcal{U}(0, 1) \) and let the observations be,

\[ h(x_k) = \begin{cases} 
    x_k & \xi_i > 0.8 \\
    \beta_k x_k - 8 & \text{else}
\end{cases} \]

\( \beta_k \sim \mathcal{N}(0.5, 1/50) \).

\( h \) is applied to 7 randomly chosen variables

Remaining 13 are directly observed
EXAMPLE 1: LORENZ-96

- The result is a bimodal distribution, “cloudy/clear”
- \( \text{Obs Model Error} = \text{True Obs} - \tilde{h}(\text{True State}) \)
CORRECTING THE BIAS

- Our goal is to find \( p(b_i | y_i) \)
- We can then adjust the filter by defining a new innovation
  \[
  \hat{\epsilon}_i = \epsilon_i + \hat{b}_i = y_i - \tilde{h}(x_i^f) + \hat{b}_i
  \]
- Where \( \hat{b}_i = \mathbb{E}_{p(b_i | y_i)}[b_i] \)
- We also inflate the obs covariance by \( R_i = R^o + \hat{R}_{b_i} \)
- Where \( \hat{R}_{b_i} = \mathbb{E}_{p(b_i | y_i)}[(b_i - \hat{b}_i)(b_i - \hat{b}_i)^\top] \)
CORRECTING THE BIAS

- If we can estimate $p(b_i \mid y_i)$ we can ‘fix’ the obs
- We will use Bayes’ to find $p(b_i \mid y_i) = p(b_i)p(y_i \mid b_i)$
- We will use a simple prior $p(b_i) = \mathcal{N}(\epsilon_i, P^y_i)$
  - $\epsilon_i = y_i - h(x_i^f)$ is the innovation
  - $P^y_i$ is the innovation covariance estimate
- The real challenge is to estimate $p(y_i \mid b_i)$
- We will learn $p(y_i \mid b_i)$ from training data using the kernel estimation of conditional distributions
CORRECTING THE BIAS

- Below plots have $y_i \approx -4$
- Left is clear, right is cloudy
- Notice bimodal likelihood
LEARNING THE CONDITIONAL DISTRIBUTION

- Given training data \((y_i, b_i)\) our goal is to learn \(p(y_i \mid b_i)\)

- For a kernel \(K(\alpha, \beta) = e^{-\frac{||\alpha - \beta||^2}{\delta^2}}\) we define Hilbert spaces

\[
\mathcal{H}_y = \left\{ \sum_{i=1}^{N} a_i K(y_i, \cdot) : \bar{a} \in \mathbb{R}^N \right\}, \mathcal{H}_b = \left\{ \sum_{i=1}^{N} a_i K(b_i, \cdot) : \bar{a} \in \mathbb{R}^N \right\}
\]

- For example the kernel density estimate (KDE) \(\hat{q}\) is in \(\mathcal{H}_y\)

\[
\hat{q}(y) = \frac{1}{m_0 N} \sum_{i=1}^{N} K(y_i, y)
\]

- Eigenvectors \(\phi_\ell\) of \(K_{ij} = K(y_i, y_j)\) form an orthonormal basis for \(\mathcal{H}_y\). Similarly \(\varphi_k\) are a basis for \(\mathcal{H}_b\).
LEARNING THE CONDITIONAL DISTRIBUTION

- We assume that $p(y | b)$ can be approximated in $\mathcal{H}_y \otimes \mathcal{H}_b$

- Let $C_{ij}^{yb} = \langle \phi_i, \varphi_j \rangle$ and $C_{ij}^{bb} = \langle \varphi_i, \varphi_j \rangle$ then define

$$C^{y|b} = C^{yb} \left( C^{bb} + \lambda I \right)^{-1}$$

- We can then define a consistent estimator of $p(y | b)$ by

$$\hat{p}(y | b) = \sum_{i,j=1}^{N} C_{i,j}^{y|b} \phi_i(y) \varphi_j(b) \hat{q}(y)$$

- We define eigenfunctions with Nystöm extension

$$\varphi_j(b) = \lambda_j^{-1} \sum_{i=1}^{N} \varphi_j(b_i) K(b_i, b)$$
OVERVIEW

- **Learning Phase:** Given training data set \((x_i, y_i)\)
  - Compute the biases \(b_i = y_i - \tilde{h}(x_i)\)
  - Learn the conditional distribution \(p(y | b)\)

- **Filtering:** Forecast \(x_i^f \Rightarrow\) innovation \(\epsilon_i = y_i - \tilde{h}(x_i^f)\)
  - Use prior \(p(b) = \mathcal{N}(\epsilon_i, P_i^y)\)
  - Combine with conditional to find \(p(b | y_i) = p(b)p(y_i | b)\)
  - Estimate conditional mean \(\hat{b}_i\) and covariance \(\hat{R}_{bi}\)
  - Adjust innovation \(\hat{\epsilon}_i = \epsilon_i + \hat{b}_i\) and \(R_i = R^o + \hat{R}_{bi}\)
  - Apply Kalman update, continue to the next filter step
OVERVIEW

Prior $p(x_i)$ $\xrightarrow{\text{Primary Filter}}$ Posterior $p(x_i | y_i)$

Error Prior $p(b)$ $\xrightarrow{\text{Secondary Filter}}$ Error Posterior $p(b | y_i)$

Observation $y_i$ $\xrightarrow{\text{RKHS+Training Data}}$ Likelihood $p(y_i | b)$
LORENZ-96 RESULTS
Lorenz-96 Results

- Works well with small measurement noise
- Observations need to be precise, but not accurate
EXAMPLE 2: MULTI-CLOUD “SATELLITE-LIKE” OBS

- Consider a 7-dim’l model for a column of atmosphere
  - Baroclinic anomaly potential temperatures, $\theta_1$ and $\theta_2$
  - Boundary layer anomaly potential temperature, $\theta_{eb}$
  - Vertically averaged water vapor content, $q$
  - Cloud fractions: congestus $f_c$, deep $f_d$, and stratiform $f_s$
- Extrapolate anomaly potential temperature at height $z$

\[
T(z) = \theta_1 \sin\left(\frac{Z\pi}{Z_T}\right) + 2\theta_2 \sin\left(\frac{2Z\pi}{Z_T}\right), \quad z \in [0, 16]
\]

EXAMPLE 2: MULTI-CLOUD “SATELLITE-LIKE” OBS

▶ Extrapolate anomaly potential temperature at height $z$

$$T(z) = \theta_1 \sin\left(\frac{Z_T}{Z_T}\right) + 2\theta_2 \sin\left(\frac{2Z_T}{Z_T}\right), \quad z \in [0, 16]$$

▶ Brightness temperature-like quantity at wavenumber-$\nu$

$$h_\nu(x, f) = (1 - f_d - f_s)\left[(1 - f_c)(\theta_{eb} T_\nu(0) + \int_0^{z_c} T(z) \frac{\partial T_\nu}{\partial z}(z) \, dz) \right.$$  
$$+ f_c T(z_c) T_\nu(z_c) + \int_{z_c}^{z_d} T(z) \frac{\partial T_\nu}{\partial z}(z) \, dz \bigg] \tag{1}$$ 
$$+ (f_d + f_s) T(z_d) T_\nu(z_d) + \int_{z_d}^{\infty} T(z) \frac{\partial T_\nu}{\partial z}(z) \, dz,$$

▶ Setting $f = 0$ is the clear sky model
EXAMPLE 2: MULTI-CLOUD “SATellite-LIKE” OBS

- Weighting functions define RTM at different wavenumbers
EXAMPLE 2: MULTI-CLOUD “SATELLITE-LIKE” OBS

- Biases at the 16 observed wavenumbers
EXAMPLE 2: MULTI-CLOUD “SATELLITE-LIKE” OBS

- Multimodal likelihood functions

![Graphs showing multimodal likelihood functions with probability distributions for different model errors.](image)
EXAMPLE 2: MULTI-CLOUD “SATELLITE-LIKE” OBS

![Graphs showing multimodal distributions with legends for Truth, EnKF, True Obs, EnKF, Wrong Obs, and RKHS. Each graph plots time series data for different parameters against time.](image-url)
EXAMPLE 2: MULTI-CLOUD "SATELLITE-LIKE" OBS
Example 2: Multi-cloud “Satellite-like” Obs

- **θ₁, MSE** (percent of variance)
- **θ₂, MSE** (percent of variance)
- **θₑₑ, MSE** (percent of variance)
- **q, MSE** (percent of variance)

Graphs showing the effect of measurement noise on MSE for different methods: EnKF, True Obs, EnKF, Wrong Obs, and RKHS.
EXAMPLE 2: MULTI-CLOUD “SATELLITE-LIKE” OBS

[Graphs showing measurement noise vs. MSE for different methods (EnKF, True Obs, EnKF, Wrong Obs, RKHS) across varying levels of measurement noise. The graphs illustrate the performance and MSE of EnKF with and without correct observations and the RKHS method under different noise conditions.]